

# Weekly Report

Lu Junhua

2016 年 6 月 6 日

## Done

- Finished code assignments of applied maths in CS, wrote a report for all the algorithms and implementations; translated a CVPR paper, which elaborates a  $k$ -means algorithm that utilizes *triangle inequality* to reduce distance calculation.
- Split the patent into different parts. The *model* and *visualization* part were finished. The *workflow* part or *system architecture* part will be done next week.
- Read some papers and browse for materials regarding game data.

## To do

- Begin review the VINCI paper.
- Finish another patent.
- Read five more papers.
- Write essays and prepare for final exams of several courses.

论文 看的论文主要偏向算法那边, 而不是可视化. 但我了解到天野写了一篇网络异常检测可视化的论文, 我会下周问问她一些情况.

- 找到一份中文的游戏运营数据分析指标白皮书. 里面提供了常见并且规范化的指标: 用户获取(acquisition), 用户活跃(activation), 留存与流失(retention & churn)和游戏收入(revenue), 每个条目下面还有更细一层的多种指标以及各自的计算方法. 在15 年的文章中, 对前三个大指标其实都是做了统计分析的. 在我近来找到的一些文章和资料中, 做churn prediction的比较常见, 因为这个问题不仅可以在游戏中做, 在其他产品或者服务也有相关研究. 我会据此再找找国内外的资料, 看看是否有对其他指标进行分析的.
- 两篇综述:

第一篇是被引用次数非常高的Anomaly Detection : A Survey (2009), 长达70页详细概述了异常检测的各种方法. 我看的是其ppt, 从大的角度来说, anomaly detection主要分为点异常检测以及其他上下文异常检测, 共有异常检测, 以及在线异常检测和分布式异常检测. 在点异常检测中, 可以看到大量传统的机器学习方法, 但是都针对异常的特性做出了改进, 考虑到异常往往数量比较小, 或者其周围分布也有一定的特性.

第二篇是比较2015年的A Survey and Analysis of Techniques for Player Behavior Prediction in Massively Multiplayer Online Role-Playing Games. 文章首先定义了一些desiderata, 作为评估某种技术的指标. 随后列举了三种玩家(行为)建模技术manual tagging, collaborative filtering和goal recognition, 按照这些指标进行分析. 这三种方法也不是之前常见的方法, 本文提供了不少相关的文献, 后面会进行评估.

- 两篇FPS(第一人称射击游戏)中基于行为的外挂检测方法, (FPS游戏也属于MMORPG, 但是在主要的活动内容上仍与倩女幽魂有所区别; 此外他们用的数据, 都是自己模拟的(我开着/不用外挂去让你检测), 和实际情况仍有一定区别)

13年Behavioral-Based Cheating Detection in Online First Person Shooters using Machine Learning Techniques. 本文亮点在于其特征的选取, 是基于经验的构造了18种(而不是基于数据挖掘方法). 这是需要对游戏的更深入的了解的. 数据按时间变化提取以后, 用分类器来进行预测. 分类器用的只是一般分类器.

06年Detecting Cheaters for Multiplayer Games: Theory, Design and Implementation. 本文用的是动态贝叶斯网络方法, 给我们的研究后面提供了新视角. 同样也是有基于经验的一些变量作为贝叶斯网中的节点.

- 09年Churn Prediction in MMORPGs: A Social Influence Based Approach. 这篇文章从social influence角度来建立网络中的diffusion model. 其实在今年投稿之前, 我们也有类似想法, 比如每个人对其他人的影响有积极的(让你留下来), 有消极的(让你别玩了). 但是我们当时认为, 由于我们没有文本所以无法语义推断消息的积极还是消极. 然而这篇文章的这个方法就是一个不需要文本介入, 而是在一定假设下推断模型参数, 用于预测churn. 这个方法如果可以结合可视化, 就会有点类似那种经典的可视数据挖掘的套路.

我的想法是, 模型方面我还要继续多了解一些异常检测方法及其可能的可视化的方法, 因为现在两个巫老师给的方向, 都离不开可视数据挖掘的套路. 此外我后面仍要对游戏加强一下了解程度, 不管是自身体验还是咨询他人.

此外找的文章还不是很新; 倩女幽魂中的外挂, 从网上粗看可能和文章的外挂性质还不一样; 流失率和留存率如果使用时序模型, 需要对不同时间点的活动进行了解, 这个仍是问题. 后面会想一下解决这些问题的对策.